National Center for Health Statistics
2000 Data Users Conference

The New National Health Interview Survey (NHIS): An Introduction and Overview of the Future

Using SAS to Manipulate the NHIS Injury and Poisoning Data

July 26-28, 2000
Hyatt Regency Hotel
Bethesda, Maryland

Centers for Disease Control and Prevention
National Center for Health Statistics
Division of Health Interview Statistics

Table of Contents

NOTE: *Example programs were run using 1997 NHIS data files. All programs in this document are presented for demonstration purposes only and are **not** intended to be "real".*

Provided SAS Program
1997 Injury Episode File
(INJEPSOD.SAS)

```
 03FEB2000:15:42:34

 THIS IS AN EXAMPLE OF A SAS PROGRAM THAT CREATES A SAS
 FILE FROM THE NHIS 1997 INJEPSOD.DAT ASCII FILE

 NOTE: THE FORMAT DEFINITIONS GIVEN BELOW WILL RESULT IN
       PROCEDURE OUTPUT SHOWING VALUES THAT HAVE BEEN
       GROUPED AS THEY ARE SHOWN IN THE FILE LAYOUT
       DOCUMENTATION

 THIS IS STORED IN INJEPSOD.SAS
*****************************************************************;
* USER NOTE: PLACE NEXT TWO STATEMENTS IN SUBSEQUENT PROGRAMS;

LIBNAME NHIS    'C:\NHIS1997\';
LIBNAME LIBRARY 'C:\NHIS1997\';

* DEFINE VARIABLE VALUES FOR REPORTS;

PROC FORMAT LIBRARY=LIBRARY;
  VALUE IEP001X
    70              = "Injury Episode"
   ;
   .
   .
   .
   .
  VALUE $IEP078X
    "E8079" - "E986"= "E codes"
    "99998"         = "Not ascertained"
    " "             = "<Blank: Not in Universe>"
   ;

DATA NHIS.INJEPSOD;
* CREATE A SAS DATA SET STORED IN DIR 'C:\NHIS1997\';
   INFILE 'C:\NHIS1997\INJEPSOD.DAT' MISSOVER TRUNCOVER LRECL=135;

   * DEFINE LENGTH OF NUMERIC VARIABLES;

   LENGTH
      RECTYPE  3   SRVY_YR  4   IJDATE_M 3   DAY      3   IJDATE_Y 4
      .
      .
      .
      LIMTM    3   IJIAD    3   HLIMT    3   WTFA     8 ;


   * INPUT ALL VARIABLES;
```

```
   INPUT
      RECTYPE       1 -   2    SRVY_YR       3 -   6
      .
      .
      .
      .
      ECODE_3  $  125 - 129    WTFA        130 - 135;

   * DEFINE VARIABLE LABELS;

   LABEL
      RECTYPE  ="Record Type"
      .
      .
      .
      .
      WTFA     ="Weight - Final Annual"
   ;

   * ASSOCIATE VARIABLES WITH FORMAT VALUES;

   FORMAT
      RECTYPE  IEP001X.    SRVY_YR  IEP002X.    INJEPNO  $IEP006X.
      .
      .
      .
      .
      ECODE_3  $IEP078X.;

PROC CONTENTS DATA=NHIS.INJEPSOD;

PROC FREQ DATA=NHIS.INJEPSOD;
   TITLE1 'FREQUENCY REPORT FOR 1997 NHIS INJEPSOD FILE';
   TITLE2 '(WEIGHTED)';
   TABLES RECTYPE/LIST MISSING;
   WEIGHT WTFA ;

PROC FREQ DATA=NHIS.INJEPSOD;
   TITLE1 'FREQUENCY REPORT FOR 1997 NHIS INJEPSOD FILE';
   TITLE2 '(UNWEIGHTED)';
   TABLES RECTYPE/LIST MISSING;
* USER NOTE: TO SEE UNFORMATTED VALUES IN PROCEDURES, ADD THE
       STATEMENT: FORMAT _ALL_;
RUN;
```

Parts of the above SAS program have been cut out so the program would fit in this handout. The entire SAS program can be viewed on:

CD Rom
Website - www.cdc.gov/nchs/nhis.htm

Example of All Variables Formatted with Value Labels
SAS Program and SAS Output
1997 Injury Episode File

SAS Program:

```
proc freq data=nhis.injepsod;
  table caus icd9_2 / missing;
title;
run;
```

SAS Output:

Cause of injury

CAUS	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Vehicle as transportation (see Notes below)	411	14.0	411	14.0
Gun/being shot	5	0.2	416	14.2
Fire/burn/scald related	46	1.6	462	15.7
Near drowning/water in lungs	3	0.1	465	15.8
Fall	1030	35.1	1495	50.9
Other	1430	48.7	2925	99.7
Refused	3	0.1	2928	99.8
Don't know	7	0.2	2935	100.0

ICD-9-CM diagnosis code

ICD9_2	Frequency	Percent	Cumulative Frequency	Cumulative Percent
<Blank: Not in Universe>	2507	85.4	2507	85.4
ICD-9-CM codes	426	14.5	2933	99.9
Not ascertained	2	0.1	2935	100.0

In this program:

The "MISSING" option in the table statement causes SAS to include missing values in calculations of frequencies and percentages.

The "TITLE" statement followed by a semi-colon prevents SAS from printing any page titles on the output generated from this program.

Example of Value Label Formatting Removed for One Variable SAS Program and SAS Output 1997 Injury Episode File

SAS Program:

```
proc freq data=nhis.injepsod;
  table caus icd9_2 / missing;
format icd9_2;
title;
run;
```

SAS Output:

Cause of injury

CAUS	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Vehicle as transportation (see Notes below)	411	14.0	411	14.0
Gun/being shot	5	0.2	416	14.2
Fire/burn/scald related	46	1.6	462	15.7
Near drowning/water in lungs	3	0.1	465	15.8
Fall	1030	35.1	1495	50.9
Other	1430	48.7	2925	99.7
Refused	3	0.1	2928	99.8
Don't know	7	0.2	2935	100.0

ICD-9-CM diagnosis code

ICD9_2	Frequency	Percent	Cumulative Frequency	Cumulative Percent
	2507	85.4	2507	85.4
80500	2	0.1	2509	85.5
80700	1	0.0	2510	85.5
80709	1	0.0	2511	85.6
.				
.				
.				
.				
9597	8	0.3	2929	99.8
9598	2	0.1	2931	99.9
9599	2	0.1	2933	99.9
99998	2	0.1	2935	100.0

In this program:

The "FORMAT" statement removes the previously assigned formatting for variable ICD9_2.

Concatenating Files (Episodes)
Creating Annual Weight Variable
1997 Injury Episode File and 1997 Poisoning Episode File

SAS Program:

```
data injpoi;
set nhis.injepsod
    nhis.poiepsod (rename=(pday=day poidtem=ijdate_m poidtey=ijdate_y
      rpckdmp=rpckdm phosp=ihosp phno=ihno pwkls=wkls pscls=scls));
wt=wtfa*4;
run;
```

SAS Log:

```
1   data injpoi;
2  set nhis.injepsod
3       nhis.poiepsod (rename=(pday=day poidtem=ijdate_m poidtey=ijdate_y
4             rpckdmp=rpckdm phosp=ihosp phno=ihno pwkls=wkls pscls=scls));
5  wt=wtfa*4;
6  run;

NOTE: The data set WORK.INJPOI has 3161 observations and 84 variables.
NOTE: The DATA statement used 3.12 seconds.
```

In this program:

The "RENAME" statement assigns new names to variables that already exist.

WT is a new variable that is created by taking variable WTFA and multiplying it by four.

Merging Files (Episodes)
1997 Injury/Poisoning Episode File and 1997 Person File

SAS Program:

```
proc sort data=injpoi;
  by hhx fmx px;
run;

proc sort data=nhis.personsx;
  by hhx fmx px;
run;

data ipp;
merge injpoi
      nhis.personsx (keep=hhx fmx px age_p sex racerec psu stratum wtfa);
  by hhx fmx px;
run;
```

SAS Log:

```
1   proc sort data=injpoi;
2     by hhx fmx px;
3   run;

NOTE: The data set WORK.INJPOI has 3161 observations and 84 variables.
NOTE: The PROCEDURE SORT used 0.11 seconds.

4   proc sort data=nhis.personsx;
5     by hhx fmx px;
6   run;

NOTE: The data set NHIS.PERSONSX has 103477 observations and 525 variables.
NOTE: The PROCEDURE SORT used 0.17 seconds.

7   data ipp;
8   merge injpoi
9         nhis.personsx (keep=hhx fmx px age_p sex racerec psu stratum wtfa);
10    by hhx fmx px;
11  run;

NOTE: The data set WORK.IPP has 103652 observations and 89 variables.
NOTE: The DATA statement used 1 minute 9.48 seconds.
```

In this program:

The "KEEP" statement causes SAS to write only those variables that are listed in the "KEEP" statement to the new data set. When there is no "KEEP" statement, all the variables in the data set are written to the new data set.

Merging Files (Episodes)
1997 Injury/Poisoning Episode File and 1997 Person File

Episode File (Injury and Poisoning)	Person File
Person 1 - Episode 1	Person 1 - Person Info
Person 1 - Episode 2	Person 2 - Person Info
Person 2 - Episode 1	Person 3 - Person Info
Person 4 - Episode 1	Person 4 - Person Info
Person 5 - Episode 1	Person 5 - Person Info
Person 5 - Episode 2	Person 6 - Person Info
Person 5 - Episode 3	

Merged Episode File (Injury and Poisoning) and Person File

Person 1 - Episode 1 - Person Info

Person 1 - Episode 2 - Person Info

Person 2 - Episode 1 - Person Info

Person 3 - No Episode - Person Info

Person 4 - Episode 1 - Person Info

Person 5 - Episode 1 - Person Info

Person 5 - Episode 2 - Person Info

Person 5 - Episode 3 - Person Info

Person 6 - No Episode - Person Info